

Endogenitet og eksperimenter – forskningsdesignet som løsning

Statskundskaben er fyldt med teorier om virkningen af uafhængige på afhængige variabler. Desværre er det ofte svært at fastslå, om en empirisk sammenhæng er udtryk for en kausal effekt. Uanset om den er fundet i et kvantitativt eller kvalitativt studie, kan en empirisk sammenhæng dække over mere, end at x påvirker y . Sammenhængen kan også være udtryk for, at y påvirker x . Dermed opstår et endogenitetsproblem, som indebærer, at man ikke uden videre kan udlede kausal inferens fra datamaterialet. I den situation er det relevant at overveje eksperimentet som løsning. I artiklen diskuterer vi disse forhold og giver eksempler på de væsentligste eksperimentelle designs, nemlig laboratorieeksperimentet, det naturlige eksperiment, felteksperimentet, surveyeksperimentet og kvasieksperimentet.

Hvor er videnskaben i samfundsvidenskaben?

Statskundskaben er fyldt med teorier om virkningen af uafhængige på afhængige variabler. Ifølge King, Keohane og Verba (1994: 7-9) er det et definerende træk ved *videnskabelig* statskundskabsforskning, uanset om tilgangen er kvantitativ eller kvalitativ, at målet er deskriptiv eller kausal inferens. Desværre er det ofte vanskeligt at fastslå, om en empirisk sammenhæng faktisk er udtryk for en kausal effekt. Eksempelvis har en række studier undersøgt Ostroms (1996) teori om, at samproduktion, hvor både offentligt ansatte og borgere bidrager til produktionen af offentlige ydelser, kan føre til højere effektivitet. Men det er svært at afgøre, om sammenhængen mellem samproduktion og effektivitet skyldes, at samproduktion faktisk virker – eller blot er udtryk for, at offentligt ansatte inddrager borgere, når tingene i forvejen går godt (Jakobsen og Andersen, 2013). Et andet eksempel er spørgsmålet om sammenhængen mellem størrelse og demokrati. Finifter (1970) argumenterer for, at størrelsen på en politisk enhed har en negativ effekt på borgernes politiske effektivitetsfølelse. Problemet er, at det er svært at vide, om en negativ empirisk sammenhæng er udtryk for, at Finifter har ret – eller blot skyldes, at borgere med en høj politisk effektivitetsfølelse har bosat sig i mindre enheder (Lassen og Serritzlew, 2011: 241).

Fælles for de to eksempler er, at en empirisk sammenhæng, uanset om den er fundet i et kvantitativt eller kvalitativt studie, kan dække over mere, end at x påvirker y . Sammenhængen kan også skyldes, at y påvirker x . Eller begge

dele samtidig. Med andre ord kan vi ikke regne med, at den uafhængige variabel faktisk er helt uafhængig; den påvirkes af andre faktorer i modellen og er dermed endogen. Endogenitetsproblemet kan have andre kilder, men det indebærer under alle omstændigheder, at vi ikke uden videre kan udlede kausal inferens fra datamaterialet.

Her er det relevant at overveje eksperimentet som løsning. Det har, ligesom både statistiske og kvalitative tilgange, inferens som mål. Eksperimentets styrke er, at det kan løse endogenitetsproblemet. Et eksperiment sikrer nemlig, at værdierne på den uafhængige variabel er bestemt eksogent, dvs. af forhold, der intet har at gøre med den teoretiske sammenhæng, man er interesseret i. Eksperimenter kan dog være svære at gennemføre eller finde, og kan – når det lader sig gøre – komme til at handle om nogle ret specielle empiriske cases, hvorfor resultaterne ikke uden videre kan generaliseres. Eksperimentet har altså både fordele og ulemper: Den interne validitet er høj, men den eksterne validitet kan være lav. Med andre metoder vil det ofte forholde sig lige omvendt. Et veltilrettelagt kvantitativt studie har en høj ekstern validitet, men den interne validitet er – hvis der er endogenitetsproblemer – lav. Derfor supplerer eksperimentet de traditionelle metoder godt, og derfor er det en skam, at eksperimentet er en relativt overset metode.

I artiklen fokuserer vi på endogenitetsproblemet og diskuterer, hvordan eksperimentet kan løse det. Vi giver konkrete eksempler på de væsentligste eksperimentelle designs, nemlig laboratorieeksperimentet, det naturlige eksperiment, felteksperimentet, surveyeksperimentet og kvasieksperimentet.

Traditionelle metoders akilleshæl: endogenitet

Empiriske studier i statskundskaben er fyldt med påstande om kausalitet. Det er ikke så underligt. Statskundskabens teorier handler ofte om at forklare et fænomen (y) ved at henvise til, hvordan det bestemmes af et andet fænomen (x). Formålet med et empirisk studie vil typisk være at afklare, om det faktisk forholder sig sådan, at x forårsager y . Empiriske studier har imidlertid en akilleshæl, som alt for ofte overses. For at en empirisk sammenhæng kan siges at være udtryk for en *effekt* af x på y , må x være eksogen. Men det er jo ikke sikkert, at denne antagelse holder, og det viser sig desværre, at konsekvenserne er alvorlige, uanset om man forsøger at undersøge sammenhængen med kvalitative eller kvantitative teknikker (King, Keohane og Verba, 1994: 185ff.).

Endogenitet kan måske nemmest forklares med udgangspunkt i den klassiske regressionsmodel:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad (1)$$

Endogenitetsproblemet opstår, hvis x er endogen, hvilket er tilfældet, hvis x korrelerer med e . Problemet kan opstå på flere måder, men først er det værd at minde om, at den klassiske regressionsanalyse simpelthen antager, at problemet ikke eksisterer (se eksempelvis Gujarati, 2003: 71). Antagelsen er ikke desto mindre afgørende. Endogenitetsproblemet kan have sin rod flere steder. Antonakis et al. (2010: 1090; se også Meyer, 1995) peger blandt andet på udeladte variabler, simultanitet, målefejl og common source bias. Pointen er, at mange forhold kan betyde, at x korrelerer med e , og at det derfor i ethvert empirisk studie kræver en del begrundelse, hvis man vil opretholde antagelsen om eksogenitet.

Lad os se nærmere på Antonakis et al.s (2010) første problem for at illustrere, hvad endogenitetsproblemet indebærer.¹ Antag, at vi estimerer y alene med x , men at y i virkeligheden er bestemt af både x og z , der korrelerer med hinanden, således at $z_i = \gamma_1 x_i + u_i$:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i \quad (2)$$

I stedet for at estimere (2) udelader vi en relevant variabel (nemlig z) og estimerer i stedet denne model:

$$y_i = \phi_0 + \phi_1 x_i + v_i \quad (3)$$

Konsekvenserne er alvorlige! De kan ses ved at indsætte, at $z_i = \gamma_1 x_i + u_i$ i (2):

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (\gamma_1 x_i + u_i) + e_i = \beta_0 + (\beta_1 + \beta_2 \gamma_1) x_i + \beta_2 u_i + e_i \quad (4)$$

Her er x åbenbart korreleret med fejlleddet (som er $\beta_2 (\gamma_1 x_i + u_i) + e_i$). Hvis vi tror, at vi i vores regressionsanalyse med regressionskoefficienten ϕ_1 har fundet effekten af x på y , tager vi fejl. ϕ_1 er jo *ikke* lig det, vi er interesseret i, nemlig β_1 . I stedet for at opnå et estimat af β_1 , rammer vi helt skævt med et estimat på $\beta_1 + \beta_2 \gamma_1$. Vores estimat kan være 0, samtidig med at den sande effekt, β_1 , er positiv eller negativ. Estimatet kan være positivt, mens den sande værdi er negativ. Det afhænger alene af, hvordan x hænger sammen med den udeladte variabel, z , og hvordan z påvirker y . Med andre ord: Bias fra en udeladt variabel skaber endogenitet, hvilket betyder, at vi systematisk fejlestimerer effekten af x . Heldigvis kan dette problem håndteres ved statistisk kontrol for z , forudsat naturligvis at z er observerbar. Det er derfor, at det er så centralt at kontrollere for relevante tredjevariabler.

En anden kilde til endogenitet er simultanitet, nemlig at x og y påvirker hinanden gensidigt. Simultanitet er et slemt problem i dobbelt forstand. For det første er det – som vi skal vende tilbage til i næste afsnit – et meget udbredt problem, særligt indenfor samfundsvidenskaberne. For det andet er det, som vi ser på i et senere afsnit, meget vanskeligt og ofte umuligt at løse med statistiske redskaber. Her skal vi blot se, at simultanitet også leder til bias i estimatet af effekten af x . Vi tager igen udgangspunkt i (1). Hvis x er eksogen, kan vi uden videre estimere β_j . Men hvis x og y påvirker hinanden gensidigt, er det mere kompliceret. I så fald afhænger y ikke blot af x , men x afhænger også af y :

$$x_i = \gamma_j y_i + u_i \quad (5)$$

Estimeres β_j på basis af (1), altså under antagelse af, at $\gamma_j = 0$, når man, ligesom i eksemplet ovenfor, frem til et biased estimat af effekten af x (se King, Keohane og Verba, 1994: 195-196).

Diskussionen viser, at antagelsen om eksogenitet, der ligger bag traditionelle empiriske analyser, langt fra er uskyldig. Hvis antagelsen ikke er rigtig, kan vi simpelt hen ikke stole på de empiriske analyser. Problemet kan håndteres på to måder. For det første kan man forsøge at kompensere statistisk for problemet. Det kan, hvis der er tale om bias fra udeladte variable, ske ved statistisk kontrol. Skyldes problemet simultanitet, er det vanskeligere men ikke umuligt. For det andet kan man forsøge at fjerne problemets årsag. Det kan ske ved, at man ved hjælp af eksperimentel metode sikrer, at x faktisk er eksogen.

Endogenitet som særligt problem for samfundsforskningen

Samfundsforskere er i høj grad ramt af problemet. Der findes nemlig aktører i samfundet, som så at sige har til opgave at skabe endogenitetsproblemer. Det gælder fx embedsmændene i den offentlige forvaltning. De er ansat til at løse problemer. Det giver en særlig aktualitet til et bestemt aspekt af endogenitetsproblemet, nemlig gensidig kausalitet (eller simultanitet). Embedsmændene reagerer på problemer og prøver at løse dem med midler, de tror virker. Så problemerne skaber løsninger, som igen påvirker problemerne. Et konkret eksempel kan være en embedsmand i en kommune med økonomiske problemer. Embedsmanden ved, at udlicitering muligvis kan være en løsning. Derfor prøver han at overtale politikere til at forbedre den økonomiske situation ved at udlicitere. Hvis han har ret, vil kommunens økonomi forbedre sig.

Mere generelt er problemer ofte anledningen til at igangsætte nye initiativer. Eksempelvis fik det udvalg, som regeringen i 2012 bad give kommunal-reformen fra 2007 et serviceeftersyn, følgende mandat: ”Udvalget skal i dets

vurderinger lægge vægt på at understøtte en høj kvalitet i de offentlige ydelser, effektivitet, styring og tværgående prioritering, en klar ansvarsfordeling mellem myndigheder og nærhed for borgerne” (Økonomi- og Indenrigsministeriet, 2013: 5). Udvalget blev altså bedt om at pege på tiltag med bestemte effekter. Dermed skabes problemer med gensidig kausalitet for forskere, der efterfølgende vil efterprøve, om tiltagene virker.

En så tæt sammenhæng mellem problemer og tiltag er langt fra noget særsyn. Reformen er sjældent tilfældige, men er som regel en reaktion på organisatorisk performance. I den forstand er reformer endogene. Derfor er det problematisk at måle effekten af reformerne ved at se på korrelationen mellem dem og forskellige performanceindikatorer. For denne korrelation er også et produkt af den effekt, der er fra performance til reformerne. At finde eksogene reformer og organisatoriske tiltag udgør derfor en særlig udfordring.

Løsninger på endogenitetsproblemer

Vi kan altså ikke regne med, at den empiriske korrelation mellem x og y uden videre kan fortolkes som et estimat af den kausale effekt af x på y . I samfunds forskning er det endda sådan, at vi ofte kan være sikre på, at relationen mellem x og y er gensidig. Så snart der bare er en rimelig mistanke om, at x ikke blot påvirker y , men at x kunne være indført for at imødegå et problem med y , har vi at gøre med simultaneitet. Man kan tænke sig to løsninger på problemet. Den ene er at tage højde for problemet ved hjælp af statistisk metode. Den anden er at tilrettelægge designet på en måde, der sikrer, at x er eksogen. Vi ser her nærmere på den mest kendte statistiske løsning på problemet. Derefter diskuterer vi designløsningen. I næste afsnit introducerer vi eksperimentet.

Hvis vi ikke kan udelukke, at y påvirker x , hjælper statistisk kontrol os ikke. Det lader sig jo ikke gøre at kontrollere for effekten af y . Vi har brug for at få greb om den del af den uafhængige variabel, som påvirker y , men som ikke skyldes y . Det kan lade sig gøre ved hjælp af instrumentvariabler (IV). Metoden er principielt simpel: Det gælder om at finde en variabel, z , som hænger (tæt) sammen med x , men som vi kan vide med sikkerhed ikke er påvirket af y – eller mere præcist ikke korrelerer med fejleddet (se fx Angrist og Krueger, 2001).

Hariris (2012) undersøgelse af statsdannelsesprocessers betydning for udvikling af demokrati er et godt eksempel. Hariri argumenterer for, at tidlig statsdannelse hos ikke-europæiske lande har en negativ indflydelse på demokratiet, fordi tidligt udviklede stater i ringere grad er blevet påvirket af de europæiske kolonimagters parlamentariske system. Hypotesen undersøges i en almindelig regressionsanalyse af effekten på demokratiet af, hvor tidligt staten

i 111 ikke-europæiske lande er dannet. Som forventet finder Hariri en negativ sammenhæng. Jo tidligere statsdannelse, jo mindre demokrati i dag. Problemet er, at sammenhængen også kan være omvendt: Et lands styreform kan også tænkes at påvirke dets levedygtighed (Hariri, 2012: 480). Hariri løser problemet ved at identificere et fænomen, nemlig tidspunktet for landets overgang fra jæger- til jordbrugssamfundet, som korrelerer med tidspunktet for statsdannelsen (altså med den uafhængige variabel), men som næppe kan være påvirket af graden af demokrati i landet i dag (dvs. den afhængige variabel). Det er smart. Hvis det faktisk er rigtigt, at tidspunktet for overgangen hænger stærkt sammen med tidspunktet for statsdannelsen, kan variabelen bruges som IV til at forudsige tidspunktet for statsdannelsen. Disse forudsagte værdier kan umuligt være påvirket af nutidens demokrati. Hvis den forudsagte værdi for statsdannelsen påvirker demokratiet i dag, er det et stærkt indicium på, at tidspunktet for statsdannelsen faktisk har den forventede negative kausale effekt.

Den statistiske tilgang til håndteringen af endogenitetsproblemet har dog sine begrænsninger. For det første er det ret sjældent, at det er muligt at identificere en god IV. For det andet findes der ikke nogen statistisk test, der kan sikre, at en given instrumentvariabel (z) ikke påvirkes, direkte eller indirekte, af den afhængige variabel.² Blot en beskedent sammenhæng med den afhængige variabel vil kunne lede til betydelig bias, særligt hvis korrelationen mellem IV og den uafhængige variabel er lav (Bound, Jaeger og Baker, 1995). Med andre ord: IV-tilgangen er kun lejlighedsvist tilgængelig. Selv når man kan identificere en velegnet IV, findes der ikke nogen metode, der kan afgøre, om problemet er løst fuldstændigt.

Den designbaserede løsnings logik er derimod at eliminere selve kilden til endogenitetsproblemet. Hvis man i dataindsamlingen sikrer sig, at den uafhængige variabel er eksogent bestemt, kan man helt udelukke, at en empirisk sammenhæng skyldes endogenitet (Dunning, 2012: 24-25). Vi ser nu nærmere på eksperimentet som en designbaseret løsning på endogenitetsproblemet.

Eksperimenter som løsning

Et eksperiment er en særlig form for test. Baggrunden er normalt, at man ønsker at undersøge effekten af et bestemt tiltag, fx en reform eller organisatorisk ændring. I det følgende betegnes dette med begrebet intervention. Videre indebærer et eksperiment normalt, at en gruppe (eksperimentgruppen) udsættes for interventionen, mens en anden gruppe (kontrolgruppen) ikke gør. Effekten af interventionen fastlægges herefter som udviklingen i eksperimentgruppen sammenlignet med kontrolgruppen. Begreber som intervention, eksperiment-

og kontrolgrupper samt sammenligning er altså centrale begreber i eksperimentel forskning.

Skal et eksperiment defineres mere præcist, bliver det vanskeligt, for der er ikke enighed i litteraturen. Morton og Williams (2010: 42) definerer et eksperiment ved det forhold, at forskeren styrer interventionen og aktivt manipulerer den. Dunning (2012: 15-16; se også McDermott, 2013: 608) tilføjer, at inddelingen i eksperiment- og kontrolgrupper skal være randomiseret. Sekhon og Titiunik (2012: 35) lægger også vægt på randomisering, hvorimod forskerens kontrol over interventionen ikke er central. Problemet med at opstille en præcis definition er, at der findes mange typer eksperimenter, som definitionen helst skal omfatte. Men ingen af de nævnte bud dækker alle former for eksperimenter. Vi foretrækker derfor at arbejde med en mindre præcis overordnet afgrænsning af eksperimentbegrebet og så mere præcist afgrænse de forskellige typer eksperimenter.

Som overordnet afgrænsning vender vi tilbage til Cook og Campbells (1979: 5) klassiske indføring i eksperimentel analyse. De definerer et eksperiment således: "All experiments involve at least a treatment, an outcome measure, units of assignment, and some comparison from which change can be inferred and hopefully attributed to the treatment". I definitionen indgår dermed en ekstern intervention, men ikke forskerens grad af kontrol herover. Inddeling i eksperiment- og kontrolgrupper indgår også i definitionen, men ikke hvorvidt denne er randomiseret. De nævnte forhold, der ikke indgår i definitionen, kan anvendes til at sondre mellem forskellige eksperimenter. Det gøres i tabel 1.

Eksperimenterne er kategoriseret efter fire kriterier. Det første er, om der indgår eksperiment- og kontrolgrupper. Som det fremgår, gælder det alle typer eksperimenter. Der er altså tale om et generisk træk ved eksperimenter. Til gengæld kan andre undersøgelsesdesign også arbejde med eksperiment- og kontrolgrupper som fx traditionelle store-N-undersøgelser af observationsdata. Kriteriet kan altså ikke diskriminere mellem eksperimentel forskning og andre former for forskning.

Det næste kriterium er, om den eksperimentelle intervention er eksogen. Det gælder også alle former for eksperimenter. Og det er efter dette kriterium, at eksperimenter adskiller sig fra andre undersøgelsesdesigns. I ikke-eksperimentelle designs er der ingen garanti for eksogenitet, hverken i traditionelle store-N-undersøgelser eller singlecasestudier. Endogenitet og eksogenitet er sjældent absolutte størrelser, men som regel tilstede i et vist omfang. Hvis forskeren kan argumentere for, at interventionen ikke er fuldstændig eksogen, men omvendt ikke påvirket af undersøgelsesobjekterne i en sådan grad, at det påvirker resultatet, kan man tale om as-if eksogenitet.

Det tredje kriterium er, om der er sket en randomiseret inddeling i eksperiment- og kontrolgrupper. Randomisering behøver ikke være foretaget af forskeren. Nogle gange kan naturlige begivenheder eller politiske beslutninger påvirke forskellige dele af befolkningen eller dele af landet på tilfældig måde. I så tilfælde taler man om as-if randomisering (Dunning, 2012: 9-10). Hvis der ikke er sket randomisering, kan forskelle i resultatet skyldes initiale forskelle i eksperiment- og kontrolgruppen, der ikke har med interventionen at gøre. I så tilfælde må forskeren tage højde for disse forskelle ved fortolkningen af resultatet. Ved randomiseret fordeling kan den kausale effekt af interventionen derimod udledes direkte af forskellen mellem eksperiment- og kontrolgruppen. Som det fremgår af tabel 1, er randomisering ikke et generisk træk ved eksperimenter, idet kvasi eksperimenter ikke opererer med denne måde at opstille eksperiment- og kontrolgrupper.

Endelig er det fjerde kriterium, om forskeren styrer interventionen. Hvis det er tilfældet, kan den eksperimentelle variation justeres mere præcist efter forskningsspørgsmålet end i det modsatte tilfælde, hvor forskeren er nødsaget til at analysere effekten af den variation, andre har skabt. Igen er der ikke tale om et generisk træk ved eksperimenter, men om et forhold der karakteriserer bestemte eksperimenter, jf. tabel 1.

Alt i alt gælder, at den centrale forskel på eksperimentel og ikke-eksperimentel forskning ikke er et spørgsmål om, hvorvidt der anvendes eksperiment- og kontrolgrupper, hvorvidt inddelingen heri er randomiseret, eller hvorvidt forskeren kan manipulere den eksperimentelle intervention. Den afgørende skillelinje er spørgsmålet om interventionens eksogenitet. Alle eksperimenter har eksogene interventioner. Det er der ingen garanti for ved traditionelle forskningsdesigns. Derfor er alle eksperimenter velegnede til at løse endogenitetsproblemer. Dertil kommer, at flere eksperimentelle designs også kan håndtere en anden væsentlig kilde til endogenitet, nemlig udeladte variabler. I de fleste eksperimenter holdes tredjevariabler konstante. Derved kan man udelukke endogenitetsbias fra udeladte variabler.

Eksperimentelle designs

I det følgende diskuterer vi mere udførligt de fem typer eksperimenter, der er skitseret i tabel 1. For hver type giver vi en mere præcis definition, illustrerer med et konkret eksempel fra dansk forskning og diskuterer fordele og ulemper.

Laboratorieeksperimentet

Som det fremgår af tabel 1, er laboratorieeksperimenter³ et rendyrket eksperimentelt design. Det indeholder en sammenligning af eksperiment- og kontrol-

Tabel 1: Forskellige typer eksperimenter

	Sammenligning af eksperiment- og kontrolgrupper?	Interventionen er eksogen eller as-if eksogen	Inddeling i eksperiment- og kontrolgrupper er randomiseret eller as-if randomiseret?	Forskeren manipulerer interventionen?
Laboratorieeksperiment	Ja	Ja	Ja	Ja
Felteksperiment	Ja	Ja	Ja	Ja
Surveyeksperiment	Ja	Ja	Ja	Ja
Naturligt eksperiment	Ja	Ja	Ja	Nej
Kvasieksperiment	Ja	Ja	Nej	Nej
Traditionelt stort-N observationsstudie	Ja	Nej	Nej	Nej
Traditionelt singlecasestudie	Nej	Nej	Nej	Nej

grupper, interventionen er ikke blot eksogen, den er også randomiseret, og det er forskeren, der manipulerer interventionen. Dertil kommer, at eksperimentet finder sted i kontrollerede omgivelser. Derved sikres det, at ingen uvedkommende faktorer påvirker eksperimentet. Naturvidenskabelige eksperimenter kan kræve ganske avancerede laboratorier for at sikre dette; i statskundskaben er det typisk tilstrækkeligt med en dør, der kan lukkes.

Et eksempel på et dansk laboratorieeksperiment er Serritzlew (2003; se også 2005). Forskningsspørgsmålet er her, hvordan rammebudgettering påvirker væksten i udgifterne. Hypotesen er, at rammebudgettering har en effekt på væksten, men at effekten er betinget af beslutningstagernes præferencer, således at væksten bestemmes af disse to uafhængige variabler og deres interaktion. Det er klart, at en traditionel empirisk undersøgelse her står overfor et betydeligt simultanitetsproblem. Det er meget muligt, at brugen af rammebudgettering påvirker udgiftsvæksten, men det er også nærliggende, at rammebudgettering bringes i anvendelse, hvis væksten i udgifterne er høj. En eventuel sammenhæng mellem anvendelse af rammebudgettering og udgiftsvækst kan derfor ikke uden videre fortolkes som en kausal effekt af budgetlægningsmetoden.

Laboratorieeksperimentet blev designet som et budgetspil med fem deltagere, der med simpelt flertal skulle beslutte sig for et budget. Fire interventioner sikrede variation i de to uafhængige variabler og mulighed for at teste interaktionseffekten. Eftersom intet andet varierede mellem de fire interventioner, kunne forskelle i udgifterne fortolkes som kausale effekter af de uafhængige variabler. Eksperimentet viste, at rammebudgettering faktisk har den forventede betingede effekt på udgifterne.

Laboratorieeksperimentet har sine særlige fordele. For det første tillader det forskeren at styre, hvilke faktorer der påvirker deltagerne i eksperimentet. Dermed er det muligt at holde tredjevariabler konstant. For det andet sikrer randomiseringen, at enhver forskel mellem eksperiment- og kontrolgrupper kan tilskrives en effekt af interventionen. For det tredje kan data (deltagernes adfærd, beslutninger, holdning mv.) registreres med meget stor præcision. For det fjerde er det muligt i laboratorieeksperimenter at *inducere* præferencer. Ved at knytte belønninger (typisk mindre pengebeløb) til deltagernes adfærd, kan forskeren meget præcist styre deltagernes præferencer. Det er yderst vanskeligt i andre designs, eksperimentelle eller ej, og derfor anvendes laboratorieeksperimenter i særlig høj grad i undersøgelser af teorier om koordination og forhandling (se fx Roth og Kagel, 1995). Den væsentligste ulempe er, at laboratorieeksperimentet typisk har lav ekstern validitet. Det skyldes, at selve undersøgelsen foregår i et laboratorium, der jo typisk ikke ligner virkeligheden særlig godt. Dertil kommer, at deltagerne sjældent er repræsentative. Begge dele gør, at resultaterne er

bedre egnede til at teste kausalhypoteser end til at generalisere til en bestemt population.

Feltekspperimentet

Feltekspperimentet⁴ ligner i sin logik laboratorieeksperimentet. Der foregår en sammenligning af eksperiment- og kontrolgrupper, inddelingen heri er randomiseret, interventionen er eksogen, og forskeren manipulerer interventionen. Forskellen er, at feltekspperimentet, som navnet antyder, foregår i felten, dvs. hos de individer, organisationer eller steder, som den teori, der undersøges, handler om (Davenport, Gerber og Green, 2010: 69-71). Hvis hypotesen fx handler om vælgerregistrerings betydning for valgdeltagelsen, skal et feltekspperiment lave en intervention i vælgerregistreringen (som i Gosnell (1927), der ifølge Davenport, Gerber og Green (2010) gennemførte det første politologiske feltekspperiment).

Et dansk eksempel er Jakobsen og Andersen (2013), der undersøger effekten af samproduktion med et feltekspperiment. Formålet var at undersøge effekten af at inddrage forældre i sprogindlæringen blandt tosprogede børnehalebørn. Det er nærliggende at forvente, at forældre med mange ressourcer gerne vil inddrages i et sådant eksperiment. Det udgør et simultantetsproblem, som kan løses med randomiseret intervention. Forskerne opdelte derfor børnehalebørnene i en eksperiment- og kontrolgruppe, hvor børnene i eksperimentgruppen fik en sprogkuffert med hjem. Sprogkufferten indeholdt materialer, hvormed forældrene kunne læse med børnene på modersmålet. Kontrolgruppen bestod af børn, der ikke modtog en sprogkuffert. Randomiseringen betyder, at forskelle i sprogkundskaber mellem kontrol- og eksperimentgruppen kan fortolkes som en effekt af kuffertindsatsen. Studiet viser, at denne form for samproduktion har de forventede positive effekter.

Feltekspperimentet sikrer som andre eksperimentelle designs en høj intern validitet. Samtidig er den eksterne validitet ganske høj. Det skyldes, at feltekspperimentet jo netop gennemføres i den virkelige verden. Dertil kommer, at feltekspperimentet kan have en betydelig policy-relevans. Politologiske felteksp eksperimenter muliggøres ofte af, at en offentlig myndighed overvejer en ændring i praksis. Hvis ændringen designes som et feltekspperiment, vil studiet næppe kunne undgå at have praktisk relevans. Resultaterne vil netop afklare, hvilke resultater den gennemførte ændring medførte. Feltekspperimentet har også en række ulemper. For det første kan det være ganske ressourcekrævende. For det andet kan mange væsentlige politologiske forskningsspørgsmål næppe belyses med felteksp eksperimenter. Det gælder eksempelvis forhold, der er tæt reguleret (hvilket kan forhindre randomiseret intervention), eller som har en tæt politisk

bevågenhed (hvor et felteksperiment hurtigt kan blive kontroversielt). For det tredje kræver felteksperimentet ofte betydelig koordination med aktører udenfor forskningsverdenen. Sådanne aktører kan have hensyn at varetage, som gør det vanskeligt at opstille et godt forskningsdesign. For det fjerde medfører det forhold, at eksperimentet gennemføres i felten, et vist kontroltab. Det betyder, at det kan være vanskeligt at sikre, at eksperimentet gennemføres helt som designet. Fx er der fare for, at kommunikation mellem deltagere på tværs af kontrol- og eksperimentgruppe fører til, at de to grupper ikke klart kan adskilles (Sinclair, McConnell og Green, 2012). Endelig kan felteksperimentet, som påvirker borgere direkte, være etisk problematiske, særligt hvis de stiller nogle borgere dårligere end andre.

Surveyeksperimentet

Også surveyeksperimentet⁵ er et rent eksperimentelt design i den forstand, at der findes eksperiment- og kontrolgrupper, eksogen intervention, randomiseret inddeling og forskermanipulation med interventionen. Surveyeksperimentet er karakteriseret ved, at eksperimentet foregår i surveyform, dvs. ved at gruppen af respondenter tilfældigt opdeles i grupper, der udsættes for forskellige versioner af et surveyspørgsmål. Interventionen ligger altså i forskellene i spørgsmålsformuleringerne. Takket være randomiseringen kan forskellene i gruppernes svar fortolkes som effekten af interventionen.

Petersen et al. (2011) har undersøgt, hvordan holdningsdannelsen påvirkes af cues, altså stikord om politikens indhold og karakter. Hypotesen er, at borgernes holdninger til velfærdspolitik afhænger af cues om modtagerens ”fortjenstfuldhed”, således at borgerens holdning spontant påvirkes af, om modtageren er en type, der fortjener offentlig hjælp. Petersen et al. (2011) undersøger spørgsmålet ved tilfældigt at fordele respondenter i fire grupper, der hver modtager en særlig beskrivelse af en modtager af velfærdsydelser: En ung mand, en kvinde i 50’erne, en kvinde i 50’erne med en arbejdsskade og en ældre mand, der har været på arbejdsmarkedet hele sit liv. Respondenterne i de fire grupper stilles derefter det samme spørgsmål, nemlig om aktiveringskravene burde strammes. Det viser sig, at støtten til stramningen afhænger af, hvilken type modtager respondenteren er blevet præsenteret for.

Surveyeksperimentet adskiller sig ved, at det er relativt billigt og nemt at gennemføre. Randomiseringen af surveyspørgsmål kan ofte ske automatisk og en enkelt survey kan uden problemer indeholde adskillige surveyeksperimentet. En anden fordel ved surveyeksperimentet er, at det tillader et stort N. Det muliggør undersøgelser af betingede effekter, som ellers sjældent er realistiske i eksperimentelle designs. Til gengæld er det en ulempe, at interventionen i

surveyeksperimenter er lavintensiv. Hvor påvirkningen af deltagerne kan være meget betydelig i både laboratorie- og felteksperimenter, er der grænser for, hvor markant deltagerne kan påvirkes alene med formuleringsforskelle. Det betyder, at man på baggrund af surveyeksperimenter, der ikke finder en effekt af en intervention, sjældent med sikkerhed kan udelukke, at der er en effekt. Det kunne jo være, at den manglende effekt blot skyldes en svag intervention.

Naturlige eksperimenter

Naturlige eksperimenter⁶ har fået deres navn, fordi data stammer fra "naturligt" forekommende fænomener. I samfundsvidenskaberne er der dog som oftest tale om produkter af sociale eller politiske processer. Det definitorisk væsentlige er, at dataene kommer udefra, ikke fra manipulation fra forskerens side. Herved adskiller naturlige eksperimenter sig fx fra laboratorieeksperimenter. Men i lighed hermed indeholder naturlige eksperimenter randomiserede (eller as-if randomiserede) eksperiment- og kontrolgrupper. Endvidere deler naturlige eksperimenter også den egenskab med laboratorieeksperimenter, at den eksperimentelle intervention er eksogen (Dunning, 2012: 41-63).

Et eksempel på et studie baseret på et naturligt eksperiment er Lassens (2005) undersøgelse af Københavns Kommunes bydelsforsøg. Kommunen var inddelt i 15 bydele, og fire bydele blev udvalgt til forsøget. I perioden 1997-2001 administrerede demokratisk valgte bydelsråd herefter en række kommunale opgaver. I de øvrige 11 bydele stod kommunen på sædvanlig vis for de kommunale opgaver. Lassen bruger forsøget til at undersøge, om information påvirker stemmeadfærd. Han udnytter her, at kommunen i 2000 afholdt en lokal folkeafstemning i hele kommunen om, hvorvidt bydelsforsøget skulle udvides til alle bydele eller helt opgives. Efter folkeafstemningen blev et repræsentativt udsnit af vælgerne spurgt, om de deltog i afstemningen. Spørgsmålet er så, om mere informerede vælgere i højere grad deltog i afstemningen. Lassens pointe er, at bydelsforsøget udgør en eksogen påvirkning af nogle vælgeres informationsniveau, nemlig beboerne i de fire udvalgte bydele, som ved mere om bydelsforsøget end de øvrige indbyggere i kommunen. Endvidere udnytter han, at informationen er nogenlunde tilfældigt fordelt blandt vælgerne, idet de fire bydele blev udvalgt, fordi de var repræsentative for hele kommunen. Alt i alt undersøger Lassen på denne måde et eksogent og as-if randomiseret stød til visse vælgeres informationsniveau. Studiet, som viser, at information højner valgdeltagelse, er et naturligt eksperiment, fordi den eksperimentelle intervention ikke er kontrolleret fra Lassens side men skabt udefra af det politiske system.

Sammenlignet med traditionelle observationsstudier, store-N-undersøgelser og singlecasestudier har naturlige eksperimenter en bedre håndtering af endogenitetsproblemet, idet den eksperimentelle variation er eksternt betinget. De har også en bedre håndtering af kontrolproblemet, idet inddelingen i eksperiment- og kontrolgrupper er randomiseret eller as-if randomiseret. Sammenlignet med andre eksperimentelle designs har naturlige eksperimenter den fordel, at de kan belyse effekten af forhold, der er svære at manipulere af forskeren selv. Hvad betyder fx sandsynligheden for militærtjeneste for mænds politiske holdninger? Politiets patruljering for kriminalitet? Valgovervågning for valgsnyd? Det er eksempler på spørgsmål, der er belyst med naturlige eksperimenter (Erikson og Stoker, 2011; Di Tella og Shargrodsky, 2004; Hyde, 2007), og som er svære at belyse i laboratorieeksperimenter.

Ulempen ved naturlige eksperimenter er, at forskeren er afhængig af udefra kommende forhold for få adgang til data. Det indebærer et vist tilfældighedselement i, hvilke spørgsmål der kan belyses. Det betyder også, at forskeren er nødsaget til at undersøge betydningen af den variation, som det naturlige eksperiment tilbyder. Det er sjældent, at den eksperimentelle variation kan gradbøjes præcist efter forskerens teoretiske spørgsmål.

Kvasieksperimenter

Kvasieksperimenter⁷ ligner andre eksperimenter på den måde, at der findes eksperiment- og kontrolgrupper og en eksogen eksperimentel intervention. Endvidere har kvasieksperimenter det til fælles med naturlige eksperimenter, at den eksperimentelle intervention kommer udefra. Den er ikke manipuleret af forskeren, men tilvejebragt af naturen eller det politiske system. Men til forskel fra andre eksperimenter er inddelingen i eksperiment- og kontrolgrupper ikke randomiseret. Grupperne kan derfor være forskellige på mange andre parametre end deres udsættelse for den eksperimentelle intervention. Udfordringen ved kvasieksperimenter er derfor at fortolke, i hvilket omfang forskellen i udfaldet mellem grupperne skyldes initiale forskelle eller den eksperimentelle intervention (Cook og Campbell 1979: 6).

Et eksempel på et kvasieksperiment er Blom-Hansen, Houlberg og Serritzlews (under udgivelse) studie af administrative stordriftsfordele i kommunerne. Udgangspunktet er den danske kommunalreform i 2007, som indebar, at nogle kommuner blev lagt sammen, mens andre fortsatte uændret. De argumenterer for, at ændringen i sammenlægningskommunernes størrelse primært er eksternt bestemt, idet den ikke var et resultat af lokale forhold, men af regeringens reform. De har dermed en as-if eksogen intervention, en kontrol- og en eksperimentgruppe samt et mål for effektivitet både før og efter sammenlæg-

ningerne. Studiet er et kvasieksperiment, fordi kommunernes inddeling i eksperiment- og kontrolgrupper ikke er randomiseret. Hvorvidt en kommune blev lagt sammen eller fortsatte uændret var ikke tilfældigt. Fx havde regeringen ingen intention om at lægge de største kommuner sammen med andre. Derfor kan det ikke udelukkes, at forskelle i udfaldet i et vist omfang skyldes initiale forskelle snarere end interventionen. Blom-Hansen, Houlberg og Serritzlew kontrollerer derfor for en række strukturelle og økonomiske forskelle mellem kommunerne, inden de konkluderer, om sammenlægningerne har haft nogen effekt. Studiet viser, at store kommuner er betydeligt billigere i drift end små.

Sammenlignet med traditionelle observationsstudier har kvasieksperimentet en bedre håndtering af endogenitetsproblemet, idet den eksperimentelle variation er eksternt betinget. Men inden for gruppen af eksperimentelle designs er kvasieksperimentet videnskabeligt set det svageste, fordi der ikke sker en randomiseret inddeling i eksperiment- og kontrolgrupper. I modsætning til andre eksperimenter er kvasieksperimentet derfor nødt til at tage højde for initiale forskelle mellem eksperiment- og kontrolgruppen. Det er en betydelig udfordring, for man kan sjældent udelukke, at der selv efter omhyggelig indsamling af kontrolvariabler resterer ikke-observerede forskelle mellem grupperne.

Alligevel har kvasieksperimentet et stort potentiale. Videnskabeligt set er det stærkere end traditionelle observationsstudier, og uden for laboratoriet er kvasieksperimentets akilleshæl – randomiseringen – ofte et ideal, der er svært at opnå i praksis. Det er der mange grunde til (Cook og Campbell, 1979: 347-371; Green, 2010; Sinclair, McConnell og Green, 2012). Det kan eksempelvis være svært at fastholde, at kontrolgruppen ikke modtager interventionen, randomiseringsprocedurer kan være fejlbehæftede, deltagere i eksperimentgruppen kan nægte at lade sig udsætte for interventionen, eller der kan ske kommunikation mellem kontrol- og eksperimentgruppen. Samtidig gælder, at kvasieksperimentet ligesom naturlige eksperimenter kan belyse effekten af forhold, der er svære at manipulere af forskeren selv.

Laboratorieeksperimentet med randomiseret inddeling i eksperiment- og kontrolgruppe og med fuld forskerkontrol over den eksperimentelle intervention betragtes ofte som ”guldstandard” inden for videnskaben. Andre designs – selv andre eksperimentelle designs – kan højst udgøre en ”sølvstandard”. Vores diskussion af fire andre eksperimentelle designs (felteksperimenter, surveyeksperimenter, naturlige eksperimenter og kvasieksperimenter) giver et mere nuanceret billede. Alle disse eksperimentelle designs har særskilte styrker og svagheder og dermed særskilte eksistensberettigelser. Den vigtigste er, at de giver mulighed for at belyse forskningsspørgsmål, som kun meget vanskeligt lader sig underkaste laboratorieforsøg. Af samme grund hilser vi det velkom-

ment, at der ikke kun sker en stigende anvendelse af eksperimenter i samfunds-forskningen, men også indhøstes erfaringer med eksperimentelle blandingsde-signs såsom *lab-in-the-field*-eksperimenter, hvor forskeren flytter laboratoriet ud i eksperimentgruppens naturlige miljø og dermed prøver at indhøste for-delene ved både laboratorie- og felteksperimentet (Morton og Williams, 2010: 296-301), og internetbaserede eksperimenter, hvor forskeren prøver at kombi-nere laboratorie- og surveyeksperimenter (Eckel og Wilson, 2006).

Konklusion

Traditionelt fremhæves, at eksperimenters styrke ligger i muligheden for at påvise kausalsammenhænge. Selvom to variabler samvarierer, behøver sam-menhængen ikke være kausal. Med deres bedre greb om kontrol- og endogeni-tetsproblemet er eksperimenter bedre i stand til at belyse dette spørgsmål. Til gengæld er der ofte noget kunstigt over eksperimenter, hvorfor man ikke altid uden videre kan overføre resultaterne fra den stiliserede eksperimentsituation til virkelighedens mangefacetterede verden.

Sagen kan også fremstilles som et spørgsmål om forholdet mellem intern og ekstern validitet (Cook og Campbell, 1979: 37-94; Serritzlew, 2007; Morton og Williams, 2010: 253-277). Intern validitet handler om, i hvilket omfang forskeren kan fastslå, om sammenhængen mellem to faktorer er kausal. Den eksterne validitet handler om, i hvilket omfang resultaterne har gyldighed ud over den kontekst, hvori undersøgelsen er foretaget. Efter denne målestok lig-ger eksperimenters styrke i den interne validitet, mens ikke-eksperimentelle metoder kan have højere ekstern validitet. Hermed mener vi også at have sagt noget om eksperimenters nødvendighed i samfundsvidenskabelig forskning. Givet eksperimenters begrænsede udbredelse kan forskningen siges i praksis at have privilegeret den eksterne validitet på bekostning af den interne. Efter vores opfattelse er dette valg svært at retfærdiggøre. En mere balanceret hen-syntagen til intern og ekstern validitet kræver mere plads til eksperimenter. Dette gælder ikke mindst, fordi kritikken af manglende eksterne validitet i eksperimentel forskning mest er rettet mod laboratorieforsøg. Andre eksperimen-telle design, måske især felteksperimenter og naturlige eksperimenter, har ret høj ekstern validitet. Det er alt i alt svært at finde på gode argumenter for at bremse udbredelsen af eksperimentel statskundskabsforskning.

Noter

1. Den følgende formelle fremstilling bygger på Antonakis et al. (2010: 1091).
2. Kriteriet er, om z er korreleret med fejlleddet, og da fejlleddet er uobserveret, kan vi ikke kende korrelationen (Wooldridge, 2013: 492).

3. Se introduktion i Iyengar (2011) og på dansk Serritzlew (2007).
4. Davenport, Gerber og Green (2010) og Gerber (2011) er begge gode introduktioner.
5. Sniderman (2011) introducerer surveyeksperimentet. *Politica* 39 (1), 2007 indeholder en kort introduktion til surveyeksperimentet og gode eksempler på konkrete studier.
6. Dunning (2012) er en god introduktion.
7. Cook og Campbell (1979) giver en klassisk introduktion.

Litteratur

- Angrist, Joshua og Alan B. Krueger (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* 15 (4): 69-85.
- Antonakis, John, Samuel Bendahan, Philippe Jacquart og Rafael Lalive (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly* 21 (6): 1086-1120.
- Blom-Hansen, Jens, Kurt Houllberg og Søren Serritzlew (under udgivelse). Size, democracy, and the economic costs of running the political system. *American Journal of Political Science*.
- Bound, John, David A. Jaeger og Regina M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90 (430): 443-450.
- Cook, Thomas D. og Donald T. Campbell (1979). *Quasi-Experimentation. Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Davenport, Tiffany C., Alan S. Gerber og Donald P. Green (2010). Field experiments and the study of political behavior, pp. 69-88 i Jan E. Leighley (red.), *The Oxford Handbook of American Elections and Political Behavior*. Oxford: Oxford University Press.
- Di Tella, Rafael og Ernesto Schargrodsky (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review* 94: 115-133.
- Dunning, Thad (2012). *Natural Experiments in the Social Sciences. A Design-Based Approach*. Cambridge: Cambridge University Press.
- Eckel, Cathrine C. og Rick K. Wilson (2006). Internet cautions: Experimental games with internet partners. *Experimental Economics* 9: 53-66.
- Erikson, Robert og Laura Stoker (2011). Caught in the draft: The effects of Vietnam draft lottery status on political attitudes. *American Political Science Review* 105: 221-237.

- Finifter, Ada W. (1970). Dimensions of political alienation. *American Political Science Review* 64 (2): 389-410.
- Gerber, Alan S. (2011). Field experiments in political science, kapitel 9 i James N. Druckman, Donald P. Green, James H. Kuklinski og Arthur Lupia (red.), *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.
- Gosnell, Harold F. (1927). *Getting-Out-the-Vote: An Experiment in the Simulation of Voting*. Chicago: Chicago University Press.
- Green, Jane (2010). Points of intersection between randomized experiments and quasi-experiments. *Annals of the American Academy of Political and Social Sciences* 628: 97-111.
- Gujarati, Damodar N. (2003). *Basic Econometrics*. Boston: McGraw Hill.
- Hariri, Jacob Gerner (2012). The autocratic legacy of early statehood. *American Political Science Review* 106 (3): 471-494.
- Hyde, Susan D. (2010). The observer effect in international politics: Evidence from a natural experiment. *World Politics* 60: 37-63.
- Iyengar, Shanto (2011). Laboratory experiments in political science, kapitel 6 i James N. Druckman, Donald P. Green, James H. Kuklinski og Arthur Lupia (red.), *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.
- Jakobsen, Morten og Simon Calmar Andersen (2013). Coproduction and equity in public service delivery. *Public Administration Review* 73 (5): 704-713.
- Kagel, John H. og Alvin E. Roth (1995). *Handbook of Experimental Economics*. Princeton: Princeton University Press.
- King, Gary, Robert O. Keohane og Sidney Verba (1994). *Designing Social Inquiry*. Princeton: Princeton University Press.
- Lassen, David Dreyer (2005). The effect of information on voter turnout: Evidence from a natural experiment. *American Journal of Political Science* 49: 103-118.
- Lassen, David Dreyer og Søren Serritzlew (2011). Jurisdiction size and local democracy: Evidence on internal political efficacy from large-scale municipal reform. *American Political Science Review* 105 (2): 238-259.
- McDermott, Rose (2013). The ten commandments of experiments. *PS: Political Science and Politics* 46 (3): 605-611.
- Meyer, Bruce D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics* 13 (2): 151-161.
- Morton, Rebecca og Kenneth C. Williams (2010). *Experimental Political Science and the Study of Causality. From Nature to the Lab*. Cambridge: Cambridge University Press.

- Ostrom, Elinor (1996). Crossing the great divide: Coproduction, synergy, and development. *World Development* 24 (6): 1073-1087.
- Petersen, Michael Bang, Rune Slothuus, Rune Stubager og Lise Tøgeby (2011). Deservingness versus values in public opinion on welfare: The automaticity of the deservingness heuristic. *European Journal of Political Research* 50: 24-52.
- Sekhon, Jasjeet S. og Rocio Titiunik (2012). When Natural Experiments are Neither Natural nor Experiments. *American Political Science Review* 106: 35-57.
- Serritzlew, Søren (2003). Kan udgiftsrammer begrænse væksten i budgetterne? *Politica* 35 (3): 255-273.
- Serritzlew, Søren (2005). The perverse effect of spending caps. *Journal of Theoretical Politics* 17 (1): 75-105.
- Serritzlew, Søren (2007). Det politologiske eksperiment. *Politica* 39 (3): 275-294.
- Sinclair, Betsy, Margaret McConnell og Donald P. Green. (2012). Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science* 56: 1055-1069.
- Sniderman, Paul M. (2011). The logic and design of the survey experiment: An autobiography of a methodological innovation, kapitel 8 i James N. Druckman, Donald P. Green, James H. Kuklinski og Arthur Lupia (red.), *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.
- Wooldridge, Jeffrey M. (2013). *Introductory Econometrics. A Modern Approach*. South-Western Cengage Learning.
- Økonomi- og Indenrigsministeriet (2013). *Evaluerings af kommunalreformen. Afrapportering fra udvalget om evaluering af kommunalreformen*. København.